Title
How to Establish Clinical Prediction Models

Authors
Lee, Yong-ho
Bang, Heejung
Kim, Dae Jung

Peer reviewed

How to Establish Clinical Prediction Models

Yong-ho Lee[1], Heejung Bang[2], Dae Jung Kim[3]

[1]Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Korea; [2]Division of Biostatistics, Department of Public Health Sciences, University of California Davis School of Medicine, Davis, CA, USA; [3]Department of Endocrinology and Metabolism, Ajou University School of Medicine, Suwon, Korea

A clinical prediction model can be applied to several challenging clinical scenarios: screening high-risk individuals for asymptomatic disease, predicting future events such as disease or death, and assisting medical decision-making and health education. Despite the impact of clinical prediction models on practice, prediction modeling is a complex process requiring careful statistical analyses and sound clinical judgement. Although there is no definite consensus on the best methodology for model development and validation, a few recommendations and checklists have been proposed. In this review, we summarize five steps for developing and validating a clinical prediction model: preparation for establishing clinical prediction models; dataset selection; handling variables; model generation; and model evaluation and validation. We also review several studies that detail methods for developing clinical prediction models with comparable examples from real practice. After model development and vigorous validation in relevant settings, possibly with evaluation of utility/usability and fine-tuning, good models can be ready for the use in practice. We anticipate that this framework will revitalize the use of predictive or prognostic research in endocrinology, leading to active applications in real clinical practice.

Keywords: Clinical prediction model; Development; Validation; Clinical usefulness

INTRODUCTION

Hippocrates emphasized prognosis as a principal component of medicine [1]. Nevertheless, current medical investigation mostly focuses on etiological and therapeutic research, rather than prognostic methods such as the development of clinical prediction models. Numerous studies have investigated whether a single variable (e.g., biomarkers or novel clinicobiochemical parameters) can predict or is associated with certain outcomes, whereas establishing clinical prediction models by incorporating multiple variables is rather complicated, as it requires a multi-step and multivariable/multifactorial approach to design and analysis [1].

Clinical prediction models can inform patients and their physicians or other healthcare providers of the patient's probability of having or developing a certain disease and help them with associated decision-making (e.g., facilitating patient-doctor communication based on more objective information). Ap-

plying a model to a real world problem can help with detection or screening in undiagnosed high-risk subjects, which improves the ability to prevent developing diseases with early interventions. Furthermore, in some instances, certain models can predict the possibility of having future disease or provide a prognosis for disease (e.g., complication or mortality). This review will concisely describe how to establish clinical prediction models, including the principles and processes for conducting multivariable prognostic studies and developing and validating clinical prediction models.

CONCEPT OF CLINICAL PREDICTION MODELS

In the era of personalized medicine, prediction of prevalent or incident diseases (diagnosis) or outcomes for future disease course (prognosis) became more important for patient management by health-care personnel. Clinical prediction models are used to investigate the relationship between future or unknown outcomes (endpoints) and baseline health states (starting point) among people with specific conditions [2]. They generally combine multiple parameters to provide insight into the relative impacts of individual predictors in the model. Evidence-based medicine requires the strongest scientific evidence, including findings from randomized controlled trials, meta-analyses, and systematic reviews [3]. Although clinical prediction models are partly based on evidence-based medicine, the user must also adopt practicality and an artistic approach to establish clinically relevant and meaningful models for targeted users.

Models should predict specific events accurately and be relatively simple and easy to use. If a prediction model provides inaccurate estimates of future-event occurrences, it will mislead healthcare professionals to provide insufficient management of patients or resources. On the other hand, if a model has high predictability power but is difficult to apply (e.g., with complicated calculation or unfamiliar question/item or unit), time consuming, costly [4] or less relevant (e.g., European model for Koreans, event too far away), it will not be commonly used. For example, a diabetes prediction model developed by Lim et al. [5] has a relatively high area under the receiver operating curve (AUC, 0.77), while blood tests that measure hemoglobin A1c, high density lipoprotein cholesterol, and triglyceride are included in the risk score, which would generally require clinician's involvement so could be a major barrier for use in community settings. When prediction models consist of complicated mathematical equations [6,7], a web-based application can

enhance implementation (e.g., calculating 10-year and lifetime risk for atherosclerotic cardiovascular disease [CVD] is available at http://tools.acc.org/ASCVD-Risk-Estimator/). Therefore, achieving a balance between predictability and simplicity is a key to a good clinical prediction model.

STEPS TO DEVELOPING CLINICAL PREDICTION MODELS

There are several reports [1,8-13] and a textbook [14] that detail methods to develop clinical prediction models. Although there is currently no consensus on the ideal construction method for prediction models, the Prognosis Research Strategy (PROGRESS) group has proposed a number of methods to improve the quality and impact of model development [2,15]. Recently, investigators on the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) study have established a checklist of recommendations for reporting on prediction or prognostic models [16]. This review will summarize the analytic process for developing clinical prediction models into five stages.

Stage 1: preparation for establishing clinical prediction models

The aim of prediction modeling is to develop an accurate and useful clinical prediction model with multiple variables using comprehensive datasets. First, we have to articulate several important research questions that affect database selection and the approach of model generation. (1) What is the target outcome (event or disease) to predict (e.g., diabetes, CVD, or fracture)? (2) Who is the target patient of the model (e.g., general population, elderly population ≥65 years or patients with type 2 diabetes)? (3) Who is the target user of the prediction model (e.g., layperson, doctor or health-related organization)? Depending on the answers to the above questions, researchers can choose the proper datasets for the model. The category of target users will determine the selection and handling process of multiple variables, which will affect the structure of the clinical prediction model. For example, if researchers want to make a prediction model for laypersons, a simple model with not many user-friendly questions in only a few categories (e.g., yes vs. no) could be ideal.

Stage 2: dataset selection

The dataset is one of the most important components of the clinical prediction model—often not under investigators' con-

Table 1. Characteristics of Different Clinical Prediction Models according to Their Purpose

Characteristic	Prevalent/concurrent events	Incident/future events
Data type	Cross-sectional data	Longitudinal/prospective cohort data
Application	Useful for asymptomatic diseases for screening undiagnosed cases (e.g., diabetes, CKD)	Useful for predicting the incidence of diseases (e.g., CVD, stroke, cancer)
Aim of the model	Detection	Prevention
Simplicity in model and use	More important	Less important
Example	Korean Diabetes Score [34]	ACC/AHA ASCVD risk equation [7]

CKD, chronic kidney disease; CVD, cardiovascular disease; ACC/AHA, American College of Cardiology/American Heart Association; ASCVD, atherosclerotic cardiovascular disease.

trol—and ultimately determines its quality and credibility; however, there are no general rules for assessing the quality of data [9]. Yet, there is no such thing as perfect data and prefect model. It would be reasonable to search for best-suited dataset. Oftentimes, secondary or administrative data sources must be utilized because a primary dataset with the study endpoint and all of key predictors is not available. Researchers should use different types of datasets, depending on the purpose of the prediction model. For example, a model for screening high-risk individuals with undiagnosed condition/disease can be developed using cross-sectional cohort data. However, such models may have relatively low power for predicting future incidence of disease when different risk factors come into play. Accordingly, longitudinal or prospective cohort datasets should be used for prediction models for future events (Table 1). Models for prevalent events are useful for predicting asymptomatic diseases, such as diabetes or chronic kidney disease, by screening undiagnosed cases, whereas models for incident events are useful for predicting the incidence of relatively severe diseases, such as CVD, stroke, and cancer.

A universal clinical prediction model for disease does not exist; thus, separate specific models that can individually assess the role of ethnicity, nationality, sex, or age on disease risk are warranted. For example, the Framingham coronary heart disease (CHD) risk score is generated by one of the most commonly used clinical prediction models; however, it tends to overestimate CHD risk by approximately 5-fold in Asian populations [17,18]. This indicates that models derived from one ethnicity sample may not be directly applied to populations of other ethnicities. Other specific characteristics of study populations beside ethnicity (e.g., obesity- or culture-related variables) could be important.

There is no absolute consensus on the minimal requirement for dataset sample size. Generally, large representative, contem-

porary datasets that closely reflect the characteristics of their target population are ideal for modeling and can enhance the relevance, reproducibility, and generalizability of the model. Moreover, two types of datasets are generally needed: a development dataset and a validation dataset. A clinical prediction model is first derived from analyses of the development dataset and its predictive performance should be assessed in different populations based on the validation dataset. It is highly recommended to use validation datasets from external study populations or cohorts, whenever available [19,20]; however, if it is not possible to find appropriate external datasets, an internal validation dataset can be formed by randomly splitting the original cohort into two datasets (if sample size is large) or statistical techniques such as jackknife or bootstrap resampling (if not) [21]. The splitting ratio can vary depending on the researchers' particular goals, but generally, more subjects should be allocated to the development dataset than to the validation dataset.

Stage 3: handling variables

Since cohort datasets contain more variables than can reasonably be used in a prediction model, evaluation and selection of the most predictive and sensible predictors should be done. Generally, inclusion of more than 10 variables/questions may decrease the efficiency, feasibility and convenience of prediction models, but expert's judgment that could be somewhat subjective is required to assess the need for each situation. Predictors that were previously found to be significant should normally be considered as candidate variables (e.g., family history of diabetes in diabetes risk score). It should be noted that not all significant predictors need to be included in the final model (e.g., $P<0.05$); predictor selection must be always guided by clinical relevance/judgement to prevent nonsensical or less relevant or user-unfriendly variables (e.g., socioeconomic status-related) or possible false-positive associations. Additionally,

variables which are highly correlated with others may be excluded because they contribute little unique information [22]. On the other hand, variables not statistically significant or with small effect size may still contribute to the model [23]. Depending on researcher discretion, different models that analyze different variables may be developed for targeting distinct users. For example, a simple clinical prediction model that does not require laboratory variables and a comprehensive model that does could both be designed for laypersons and health care providers, respectively [19].

With regard to variable coding, categorical and continuous variables should be managed differently [8]. For ordered categorical variables, infrequent categories can be merged and similar variables may be combined/grouped. For example, past and current smoker categories can be merged if numbers of subjects who report being a past or current smoker are relatively small and variable unification does not alter the statistical significance of the model materially. Although continuous parameters are usually included in a regression model, assuming linearity, researchers should consider the possibility of non-linear associations such as J- or U-shaped distributions [24]. Furthermore, the relative effect of a continuous variable is determined by the measurement scale used in the model [8]. For example, the impact of fasting glucose levels on the risk of CVD may be interpreted as having a stronger influence when scaled per 10 mg/dL than per 1 mg/dL.

Researchers often emphasize the importance of not dichotomizing continuous variables in the initial stage of model development because valuable predictive information can be lost during categorization [24]. However, prediction models—is not the same thing as regression models—with continuous parameters may be complex and hard to use or be understood by laypersons, because they have to calculate their risk scores by themselves. A web or computer-based platform is usually required for the implementation of these models. Otherwise, in a later phase, researchers may transform the model into a user-friendly format by categorizing some predictors, if the predictive capacity of the model is retained [8,19,25].

Finally, missing data is a chronic problem in most data analyses. Missing data can occur various reasons, including uncollected (e.g., by design), not available or not applicable, refusal by respondent, dropout, or "don't know." To handle this issue, researchers may consider imputation technique, dichotomizing the answer into yes versus others, or allow "unknown" as a separate category as in http://www.cancer.gov/bcrisktool/.

Stage 4: model generation

Although there are no consensus guidelines for choosing variables and determining structures to develop the final prediction model, various strategies with statistical tools are available [8,9]. Regression analyses, including linear, logistic, and Cox models are widely used depending on the model and its intended purpose. First, the full model approach is to include all the candidate variables in the model; the benefit of this approach is to avoid overfitting and selection bias [9]. However, it can be impractical to pre-specify all predictors and previously significant predictors may not be in a new population/sample. Second, a backward elimination approach or stepwise selection method can be applied to remove a number of insignificant candidate variables. To check for overfitting of the model, Akaike information criterion (AIC) [26], an index of model fitting that charges a penalty against larger models, may be useful [19]. Lower AIC values indicate a better model fit. Some interpret that AIC addresses explanation and Bayesian information criterion (BIC) addresses prediction, where BIC may be considered a Bayesian counterpart [27].

If researchers prefer algorithm modeling culture instead of data modeling culture, e.g., formula-based regression [28], a classification and regression tree analysis or recursive partitioning could be considered [28-30].

With regard to determining scores for each predictor in the generation of simplified models, researchers using expert judgment may create a weighted scoring system by converting β coefficients [19] or odds ratios [20] from the final model to integer values, while preserving monotonicity and simplicity. For example, from the logistic regression model built by Lee et al. [19], β coefficients <0.6, 0.7 to 1.3, 1.4 to 2.0, and >2.1 were assigned scores of 1, 2, 3, and 4, respectively.

Stage 5: model evaluation and validation (internal/external)

After model generation, researchers should evaluate the predictive power of their proposed model using an independent dataset, where truly external dataset is preferred whenever available. There are several standard performance measures that capture different aspects: two key components are calibration and discrimination [8,9,31]. Calibration can be assessed by plotting the observed proportions of events against the predicted probabilities for groups defined by ranges of individual predicted risk [9,10]. For example, a common method is to categorize 10 risk groups of equal size (deciles) and then conduct the calibration process [32]. The most ideal calibration plot would show a 45°

Table 2. Statistical Measures for Model Evaluation

Sensitivity and specificity

Discrimination (ROC/AUC)

Predictive values: positive, negative

Likelihood ratio: positive, negative

Accuracy: Youden index, Brier score

Number needed to treat or screen

Calibration: Calibration plot, Hosmer-Lemeshow test

Model determination: R^2

Statistical significance: *P* value (e.g., likelihood ratio test)

Magnitude of association, e.g., β coefficient, odds ratio

Model quality: AIC/BIC

Net reclassification index and integrated discrimination improvement

Net benefit

Cost-effectiveness

ROC, receiver operating characteristic; AUC, area under the curve; AIC, Akaike information criterion; BIC, Bayesian information criterion.

line, which indicates that the observed proportions of events and predicted probabilities completely overlap over the entire range of probabilities [9]. However, this is not guaranteed when external validation is conducted with a different sample. Discrimination is defined as the ability to distinguish events versus non-events (e.g., dead vs. alive) [8]. The most common discrimination measure is the AUC or, equivalently, concordance (c)-statistic. The AUC is equal to the probability that, given two individuals randomly selected—one who will develop an event and another who will not—the model will assign a higher probability of an event to the former [10]. A c-statistic value of 0.5 indicates a random chance (i.e., flip of a coin). The usual c-statistic range for a prediction model is 0.6 to 0.85; this range can be affected by target-event characteristics (disease) or the study population. A model with a c-statistic ranging from 0.70 to 0.80 has an adequate power of discrimination; a range of 0.80 to 0.90 is considered excellent. Table 2 shows several common statistical measures for model evaluation.

As usual, selection, application and interpretation of any statistical method and results need great care as virtually all methods entail assumptions and limited capacity. Let us review some here. Predictive values depend on the disease prevalence so direct comparison for different diseases may not be valid. When sample size is very large, *P* value can be impressively small even for a practically meaningless difference. Net reclassification index and integrated discrimination improvement are known to lead to non-proper scoring and vulnerable to miscali-

brated or overfit problems [33]. AUC and R2 are often hard to increase by a new predictor, even with large odds ratio. Despite similar names, AIC and BIC address slightly different issues and information in BIC can be decreased with sample size increases. The Hosmer-Lemeshow test is highly sensitive when sample size is large, which is not an ideal property as a goodness-fit statistic. Calibration plot can easily provide a high correlation coefficient (>0.9), simply because they are computed for predicted versus observed values on grouped data (without random variability). Finally, AUC also needs caution: a high value (e.g., >0.9) may mean excellent discrimination but it can also reflect the situation where prediction is not so relevant: (1) the task is closer to diagnostic or early onset rather than prediction; (2) cases vs. non-cases are fundamentally different with minimal overlap; or (3) predictors and endpoints are virtually the same things (e.g., current blood pressure vs. future blood pressure).

Despite the long list provided above, we do not think this is a discouraging news to researchers. We may tell us no method is perfect and "one size does not fit all" is also true to statistical methods; thus blinded or automated application can be dangerous.

It is crucial to separate internal and external validation and to conduct the previously mentioned analyses on both datasets to finalize the research findings (see the following for example reports [19,20,34]). Internal validation can be done using a random subsample or different years from the development dataset or by conducting bootstrap resampling [22]. This approach can particularly assess the stability of selected predictors, as well as prediction quality. Subsequently, external validation should be performed on an independent dataset from that which was previously used to develop the model. For example, datasets can be obtained from populations from other hospitals or centers (see geographic validation [19]) or a more recently collected cohort population (temporal validation [34]). This process is often considered to be a more powerful test for prediction models than internal validation because it evaluates transportability, generalizability and true replication, rather than reproducibility [8]. Poor model performance may occur after use of an external dataset due to differences in healthcare systems, measurement methods/definitions of predictors and/or endpoint, subject characteristics or context (e.g., high vs. low risk).

CONCLUSIONS

For patient-centered perspectives, clinical prediction models are useful for several purposes: to screen high-risk individuals

for asymptomatic disease, to predict future events of disease or death, and to assist medical decision-making. Herein, we summarized five steps for developing a clinical prediction model. Prediction models are continuously designed but few have had their predictive performance validated with an external population. Because model development is complex, consultation with statistical experts can improve the validity and quality of rigorous prediction model research. After developing the model, vigorous validation with multiple external datasets and effective dissemination to interested parties should occur before using the model in practice [35]. Web or smartphone-based applications can be good routes for advertisement and delivery of clinical prediction models to the public. For example, Korean risk models for diabetes, fatty liver, CVD, and osteoporosis are readily available at http://cmerc.yuhs.ac/mobileweb/. Simple model may be translated into a one page checklist for patient's self-assessment (e.g., equipped in waiting room in clinic). We anticipate that the framework that we provide/summarize, along with additional assistance from related references or textbooks, will help predictive or prognostic research in endocrinology; this will lead to active application of these practices in real world settings. In light of the personalized- and precision-medicine era, further research is needed to attain individual-level predictions, where genetic or novel biomarkers can play bigger roles, as well as simple generalized predictions which can further help patient-centered care.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

ORCID

Yong-ho Lee *http://orcid.org/0000-0002-6219-4942*

Dae Jung Kim *http://orcid.org/0000-0003-1025-2044*

REFERENCES

1. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009;338:b375.
2. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ 2013;346:e5595.
3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71-2.
4. Greenland S. The need for reorientation toward cost-effective prediction: comments on 'Evaluating the added predictive ability of a new marker. From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). Stat Med 2008;27:199-206.
5. Lim NK, Park SH, Choi SJ, Lee KS, Park HY. A risk score for predicting the incidence of type 2 diabetes in a middle-aged Korean cohort: the Korean genome and epidemiology study. Circ J 2012;76:1904-10.
6. Griffin SJ, Little PS, Hales CN, Kinmonth AL, Wareham NJ. Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. Diabetes Metab Res Rev 2000;16:164-71.
7. Goff DC Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014;129 (25 Suppl 2):S49-73.
8. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 2014;35:1925-31.
9. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. BMJ 2009;338:b604.
10. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338:b605.
11. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606.

12. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA 1997;277:488-94.

13. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19:453-73.

14. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer; 2009.

15. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013;10:e1001381.

16. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55-63.

17. Liu J, Hong Y, D'Agostino RB Sr, Wu Z, Wang W, Sun J, et al. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study. JAMA 2004;291:2591-9.

18. Jee SH, Jang Y, Oh DJ, Oh BH, Lee SH, Park SW, et al. A coronary heart disease prediction model: the Korean Heart Study. BMJ Open 2014;4:e005025.

19. Lee YH, Bang H, Park YM, Bae JC, Lee BW, Kang ES, et al. Non-laboratory-based self-assessment screening score for non-alcoholic fatty liver disease: development, validation and comparison with other scores. PLoS One 2014;9:e107584.

20. Bang H, Edwards AM, Bomback AS, Ballantyne CM, Brillon D, Callahan MA, et al. Development and validation of a patient self-assessment score for diabetes risk. Ann Intern Med 2009;151:775-83.

21. Kotronen A, Peltonen M, Hakkarainen A, Sevastianova K, Bergholm R, Johansson LM, et al. Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. Gastroenterology 2009;137:865-72.

22. Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.

23. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol 1996;49:907-16.

24. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006;25:127-41.

25. Boersma E, Poldermans D, Bax JJ, Steyerberg EW, Thomson IR, Banga JD, et al. Predictors of cardiac events after major vascular surgery: role of clinical characteristics, dobutamine echocardiography, and beta-blocker therapy. JAMA 2001;285:1865-73.

26. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. J R Stat Soc Ser C Appl Stat 1999;48:313-29.

27. Shmueli G. To explain or to predict? Stat Sci 2010:289-310.

28. Heikes KE, Eddy DM, Arondekar B, Schlessinger L. Diabetes risk calculator: a simple tool for detecting undiagnosed diabetes and pre-diabetes. Diabetes Care 2008;31:1040-5.

29. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. Belmont: Wadsworth International Group; 1984.

30. Breiman L. Statistical modeling: the two cultures (with comments and a rejoinder by the author). Statist Sci 2001;16:199-231.

31. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128-38.

32. Meffert PJ, Baumeister SE, Lerch MM, Mayerle J, Kratzer W, Volzke H. Development, external validation, and comparative assessment of a new diagnostic score for hepatic steatosis. Am J Gastroenterol 2014;109:1404-14.

33. Hilden J. Commentary: on NRI, IDI, and "good-looking" statistics with nothing underneath. Epidemiology 2014;25:265-7.

34. Lee YH, Bang H, Kim HC, Kim HM, Park SW, Kim DJ. A simple screening score for diabetes for the Korean population: development, validation, and comparison with other scores. Diabetes Care 2012;35:1723-30.

35. Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? BMJ 1995;311:1539.